1645

#10

PAGE: 1

RAW SEQUENCE LISTING
PATENT APPLICATION US/08/938,548B

DATE: 09/17/98
TIME: 09:12:00

INPUT SET: S28690.raw

This Raw Listing contains the General
Information Section and up to the first 5 pages

ENTERED

```
 1                              SEQUENCE LISTING
 2
 3   (1)     General Information
 4
 5           (i) APPLICANT:  Yanagisawa, Masashi
 6       Bergsma, Derk
 7       Wilson, Shelagh
 8       Brooks, David
 9       Gellai, Miklos
10
11           (ii) TITLE OF THE INVENTION:NOVEL LIGANDS OF THE NEUROPEPTIDE
12                   RECEPTOR HFGAN72
13
14           (iii) NUMBER OF SEQUENCES: 21
15
16           (iv) CORRESPONDENCE ADDRESS:
17               (A) ADDRESSEE: SmithKline Beecham Corporation
18               (B) STREET: 709 Swedeland Road
19               (C) CITY: King of Prussia
20               (D) STATE: PA
21               (E) COUNTRY: United States of America
22               (F) ZIP: 19406
23
24           (v) COMPUTER READABLE FORM:
25               (A) MEDIUM TYPE: Diskette
26               (B) COMPUTER: IBM Compatible
27               (C) OPERATING SYSTEM: DOS
28               (D) SOFTWARE: FastSEQ for Windows Version 2.0
29
30           (vi) CURRENT APPLICATION DATA:
31               (A) APPLICATION NUMBER: 08/938,548
32               (B) FILING DATE: 26-SEPT-1997
33               (C) CLASSIFICATION:
34
35   (vii) PRIOR APPLICATION DATA:
36               (A) APPLICATION NUMBER: 08/887,382
37               (B) FILING DATE: 2-JUL-1997
38
39           (vii) PRIOR APPLICATION DATA:
40               (A) APPLICATION NUMBER: 08/820,519
41               (B) FILING DATE: 19-MAR-1997
42
43               (A) APPLICATION NUMBER: 60/033,604
44               (B) FILING DATE: 17-DEC-1997
45
46           (viii) ATTORNEY/AGENT INFORMATION:
```

INPUT SET: S28690.raw

```
47              (A) NAME: Elizabeth J. Hecht
48              (B) REGISTRATION NUMBER: 41,824
49            . (C) REFERENCE/DOCKET NUMBER: ATG50037-2
50
51        (ix) TELECOMMUNICATION INFORMATION:
52             (A) TELEPHONE: 610-270-5009
53      (B) TELEFAX: 610-270-5090
54             (C) TELEX:
55
56          (2) INFORMATION FOR SEQ ID NO:1:
57
58        (i) SEQUENCE CHARACTERISTICS:
59             (A) LENGTH: 1970 base pairs
60             (B) TYPE: nucleic acid
61             (C) STRANDEDNESS: single
62           . (D) TOPOLOGY: linear
63
64        (ii) MOLECULE TYPE: Genomic DNA
65
66        (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
67
68   AAAACATAAT GTGGGTCTCG CGTCTGCCTC TCTCCCGCCC CTAATTAGCA GCTGCCTCCC        60
69   TCCATATTGT CCCAGGCCAG CGCTTCTTTT GTGCTCCCAG ATTCCTGGGT GCAAGGTGGC       120
70   CTCATTAGTG CCCGGAGACC GCCCCATCTC CAGGGAGCAG ATAGACAGAC AAGGGGGTGA       180
71   TCAGGGGCAC AGTGATCCAA CCCTGGCCTC TGAACGCCGC AGCGGCCATT CCTTGGGCCC       240
72   AGCCTGGAGA CGGCCCCCCT GCAGCAGGCT AATCTTAGAC TTGCCTTTGT CTGGCCTGGG       300
73   TGTGGACGCA ATGTGCCTGT CAATTCCCCG CCACCTCAGA GCACTATAAA CCCCAGACCC       360
74   CTGGGAGTGG GTCACAATTG ACAGCCTCAA GGTTCCTGGC TTTTTGAACC ACCACAGACA       420
75   TCTCCTTTCC CGGCTACCCC ACCCTGAGCG CCAGACACCA TGAACCTTCC TTCCACAAAG       480
76   GTAAAGATCC AGGGATGGAG GGGTGACTCA GCCATCCCAG AGGAAGCAAA AAGAGTGCTT       540
77   GCTCAGAGGG CTGGAAGAAA GGCCAAAGGT GTCTCCACTC TTGGTCTTTT CCTGGGTGTG       600
78   CTCTGAGGCA GGAGCACCTG CCTTGGCTCA CATTGGGTTG GGTGCTGTTT TGCTAAGAGC       660
79   CTGTGTTTGC TGAGCTCATA TGTGTCAGGT GCTCCGTTTG CACCTGTCAT CTCTTGTCAT       720
80   CCTCCCAACA GCCTTGCAGA GTAGAAATTA TTTCTAGTAT ACCCAGTTTA CAGGTAAGGG       780
81   AGCTGTGCCC TCTGAAAGGG CAGGAAACTG GTTCAAAGCA ACGGAGTTCA GTCACTCCTG       840
82   CAAGGGGGCA GGCAGATGAG AGAGCATTCT GGAGTCTTGC TAGTTCCTGA TTTCCATGTG       900
83   TTTCCCTGCT GTGGAGAGGA AGTTGGGGGG ACTCAGTAGG GCCCGGGTTT TTCCCAAGTT       960
84   TACAACTTCT GCTGCAGACA GACACTCCTG TTTTCAGGTG GAGTGGCAAG TGCCCTAGTG      1020
85   GTGGCAACAG TGGCCTAAGT CTCCAGAGAA AAGGGGGATT CACTCTGCCC AGGGGGTCTC      1080
86   AAAAGGCTTC CTGTGGGAGA TGCTCTGCTG GGTCTTGAAG GAGGAGCAGG GAAAGTAGGC      1140
87   CGATACCAGC AAGGGCGCAA AGCAAGGAGA ACTAAGTGAC AGCCAGAAAG GAGTGCAGGC      1200
88   TTGGAGGGGG CGCGGAGCCA GAGGGGCAGG TCCTGTGCGT GGGAGCTGGT GGCGGGCGCC      1260
89   GTGGGAAGAC CCCCCCAGCG CCCTGTCTCC GTCTCCCTAG GTCTCCTGGG CCGCCGTGAC      1320
90   GCTACTGCTG CTGCTGCTGC TGCTGCCGCC CGCGCTGTTG TCGTCCGGGG CGGCTGCACA      1380
91   GCCCCTGCCC GACTGCTGTC GTCAAAAGAC TTGCTCTTGC CGCCTCTACG AGCTGCTGCA      1440
92   CGGCGCGGGC AATCACGCGG CCGGCATCCT CACGCTGGGC AAGCGGAGGT CCGGGCCCCC      1500
93   GGGCCTCCAG GGTCGGCTGC AGCGCCTCCT GCAGGCCAGC GGCAACCACG CCGCGGGCAT      1560
94   CCTGACCATG GGCCGCCGCG CAGGCGCAGA GCCAGCGCCG CGCCCCTGCC TCGGGCGCCG      1620
95   CTGTTCCGCC CCGGCCGCCG CCTCCGTCGC GCCCGGAGGA CAGTCCGGGA TCTGAGTCGT      1680
96   TCTTCGGGCC CTGTCCTGGC CCAGGCCTCT GCCCTCTGCC CACCCAGCGT CAGCCCCCAG      1740
97   AAAAAAGGCA ATAAAGACGA GTCTCCATTC GTGTGACTGG TCTCTGTTCC TGTGCGGTCG      1800
98   CGTCCTGCCC ATCCGGGGTG GCAAAGCGTC TTGCGGAGGA CAGCTGGGCC TGGAAGCCCG      1860
99   GCTGTCGGGC ACCAGCCTTA GCTTTTGCGT GGTTGAATCG GAAACACTCT TGGTTGGGGA      1920
```

INPUT SET: S28690.raw

```
100    GTTCCCAGTG CAAGGCCCTG GGGCACAGAG AGAACTGCAC AGGTGCATGC                1970
101
102                (2) INFORMATION FOR SEQ ID NO:2:
103
104            (i) SEQUENCE CHARACTERISTICS:
105               (A) LENGTH: 131 amino acids
106               (B) TYPE: amino acid
107               (C) STRANDEDNESS: single
108               (D) TOPOLOGY: linear
109
110            (ii) MOLECULE TYPE: protein
111
112            (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
113
114    Met Asn Leu Pro Ser Thr Lys Val Ser Trp Ala Ala Val Thr Leu Leu
115     1               5                  10                  15
116    Leu Leu Leu Leu Leu Leu Pro Pro Ala Leu Leu Ser Ser Gly Ala Ala
117              20                  25                  30
118    Ala Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg
119         35                  40                  45
120
121    Leu Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu
122         50                  55                  60
123    Thr Leu Gly Lys Arg Arg Ser Gly Pro Pro Gly Leu Gln Gly Arg Leu
124    65                  70                  75                  80
125    Gln Arg Leu Leu Gln Ala Ser Gly Asn His Ala Ala Gly Ile Leu Thr
126              85                  90                  95
127    Met Gly Arg Arg Ala Gly Ala Glu Pro Ala Pro Arg Pro Cys Leu Gly
128              100                 105                 110
129    Arg Arg Cys Ser Ala Pro Ala Ala Ala Ser Val Ala Pro Gly Gly Gln
130         115                 120                 125
131    Ser Gly Ile
132         130
133
134            (2) INFORMATION FOR SEQ ID NO:3:
135
136            (i) SEQUENCE CHARACTERISTICS:
137               (A) LENGTH: 33 amino acids
138               (B) TYPE: amino acid
139               (C) STRANDEDNESS: single
140               (D) TOPOLOGY: linear
141
142            (ii) MOLECULE TYPE: protein
143
144            (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:
145
146    Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg Leu
147     1               5                  10                  15
148    Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu Thr
149              20                  25                  30
150    Leu
151
152    (2) INFORMATION FOR SEQ ID NO:4:
```

RAW SEQUENCE LISTING
PATENT APPLICATION *US/08/938,548B*

INPUT SET: S28690.raw

```
153
154            (i) SEQUENCE CHARACTERISTICS:
155              (A) LENGTH: 28 amino acids
156              (B) TYPE: amino acid
157              (C) STRANDEDNESS: single
158              (D) TOPOLOGY: linear
159
160           (ii) MOLECULE TYPE: protein
161
162           (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
163
164    Arg Ser Gly Pro Pro Gly Leu Gln Gly Arg Leu Gln Arg Leu Leu Gln
165     1               5                  10                  15
166    Ala Ser Gly Asn His Ala Ala Gly Ile Leu Thr Met
167                 20                  25
168
169            (2) INFORMATION FOR SEQ ID NO:5:
170
171            (i) SEQUENCE CHARACTERISTICS:
172              (A) LENGTH: 585 base pairs
173              (B) TYPE: nucleic acid
174              (C) STRANDEDNESS: single
175              (D) TOPOLOGY: linear
176
177           (ii) MOLECULE TYPE: cDNA
178
179           (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
180
181    GGCTCGGCGG CCTCAGACTC CTTGGGTATT TGGACCACTG CACCGAAGAT ACCATCTCTC      60
182    CGGATTGCCT CTCCCTGAGC TCCAGACACC ATGAACCTTC CTTCTACAAA GGTTCCCTGG     120
183    GCCGCCGTGA CGCTGCTGCT GCTGCTACTG CTGCCGCCGG CGCTGCTGTC GCTTGGGGTG     180
184    GACGCGCAGC CTCTGCCCGA CTGCTGTCGC CAGAAGACGT GTTCCTGCCG TCTCTACGAA     240
185    CTGTTGCACG GAGCTGGCAA CCACGCCGCG GGCATCCTCA CTCTGGGAAA GCGGCGACCT     300
186    GGACCCCCAG GCCTCCAAGG ACGGCTGCAG CGCCTCCTTC AGGCCAACGG TAACCACGCA     360
187    GCTGGCATCC TGACCATGGG CCGCCGCGCA GGCGCAGAGC TAGAGCCATA TCCCTGCCCT     420
188    GGTCGCCGCT GTCCGACTGC AACCGCCACC GCTTTAGCGC CCCGGGGCGG ATCCAGAGTC     480
189    TGAACCCGTC TTCTATCCCT GTCCTAGTCC TAACTTTCCC CTCTCCTCGC CGGTCCCTAG     540
190    GCAATAAAGA CGTTTCTCTG CTAAAAAAAA AAAAAAAAAA AAAAA                     585
191
192            (2) INFORMATION FOR SEQ ID NO:6:
193
194            (i) SEQUENCE CHARACTERISTICS:
195              (A) LENGTH: 130 amino acids
196              (B) TYPE: amino acid
197              (C) STRANDEDNESS: single
198              (D) TOPOLOGY: linear
199
200           (ii) MOLECULE TYPE: protein
201
202           (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
203
204    Met Asn Leu Pro Ser Thr Lys Val Pro Trp Ala Ala Val Thr Leu Leu
205     1               5                  10                  15
```

INPUT SET: S28690.raw

```
206   Leu Leu Leu Leu Leu Pro Pro Ala Leu Leu Ser Leu Gly Val Asp Ala
207             20                25              30
208   Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg Leu
209             35                40              45
210   Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu Thr
211       50                55              60
212   Leu Gly Lys Arg Arg Pro Gly Pro Pro Gly Leu Gln Gly Arg Leu Gln
213   65             70                75              80
214   Arg Leu Leu Gln Ala Asn Gly Asn His Ala Ala Gly Ile Leu Thr Met
215             85                90              95
216   Gly Arg Arg Ala Gly Ala Glu Leu Glu Pro Tyr Pro Cys Pro Gly Arg
217             100               105             110
218   Arg Cys Pro Thr Ala Thr Ala Thr Ala Leu Ala Pro Arg Gly Gly Ser
219       115               120             125
220   Arg Val
221       130



          (2) INFORMATION FOR SEQ ID NO:7:


          (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 32 amino acids
             (B) TYPE: amino acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear


          (ii) MOLECULE TYPE: protein


          (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:


   Met Asn Leu Pro Ser Thr Lys Val Pro Trp Ala Ala Val Thr Leu Leu
   1             5                10              15
   Leu Leu Leu Leu Leu Pro Pro Ala Leu Leu Ser Leu Gly Val Asp Ala
             20                25              30


          (2) INFORMATION FOR SEQ ID NO:8:


          (i) SEQUENCE CHARACTERISTICS:
             (A) LENGTH: 33 amino acids
             (B) TYPE: amino acid
             (C) STRANDEDNESS: single
             (D) TOPOLOGY: linear


          (ii) MOLECULE TYPE: protein


          (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:


   Gln Pro Leu Pro Asp Cys Cys Arg Gln Lys Thr Cys Ser Cys Arg Leu
   1             5                10              15
   Tyr Glu Leu Leu His Gly Ala Gly Asn His Ala Ala Gly Ile Leu Thr
             20                25              30
   Leu
```

SEQUENCE VERIFICATION REPORT
PATENT APPLICATION   *US/08/938,548B*

INPUT SET: S28690.raw

Line   Error   Original Text